

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-050499

(43)Date of publication of application : 20.02.1996

(51)Int.Cl.

G10L 9/10  
G06F 15/18  
G10L 3/00  
G10L 9/18

(21)Application number : 07-176872

(71)Applicant : AT & T CORP

(22)Date of filing : 21.06.1995

(72)Inventor : LEE CHIN-HUI  
SANKAR ANANTH

(30)Priority

Priority number : 94 263284 Priority date : 21.06.1994 Priority country : US

(54) METHOD FOR DISCRIMINATING SIGNAL

(57)Abstract:

PURPOSE: To attain sound signal recognition suitable for application in an adverse environment by sharply reducing the average error rate of sound signal recognition under a non-coincident condition.

CONSTITUTION: A set of signals (e.g. a sound signal) and stored expressions (e.g. the stored expression of a keyword) is found out and at least one conversion for generating a signal for relatively and effectively emulating the stored expression is executed (202, etc.). The operation is attained by using one of three technical elements. In 1st technique, a signal is converted so as to be more approximated (e.g. approached) by one of plural stored expressions. In 2nd technique, a set of stored expressions is converted so that a signal is more approximated by one of the stored expressions. In 3rd technique, both of the signal and the set of stored expressions are converted. In this case, HMM, a neural network and a vector quantization expression are used as stored expressions.



## LEGAL STATUS

[Date of request for examination] 05.03.1998

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3457431

[Date of registration] 01.08.2003

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision]

of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平8-50499

(43)公開日 平成8年(1996)2月20日

(51)Int.Cl. <sup>9</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 9/10	3 0 1 C			
G 0 6 F 15/18	5 6 0 G	8837-5L		
G 1 0 L 3/00	5 3 5			
9/18	E			

審査請求 未請求 請求項の数24 F D (全 19 頁)

(21)出願番号 特願平7-176872

(22)出願日 平成7年(1995)6月21日

(31)優先権主張番号 2 6 3 2 8 4

(32)優先日 1994年6月21日

(33)優先権主張国 米国 (U S)

(71)出願人 390035493

エイ・ティ・アンド・ティ・コーポレーション

AT&amp;T CORP.

アメリカ合衆国 10013-2412 ニューヨーク  
ニューヨーク アヴェニュー オブ  
ジ アメリカズ 32

(72)発明者 チン・ファイ リー

アメリカ合衆国, 07974 ニュージャージー  
ー, ニュー プロヴィデンス, ランニイメ  
イド パークウェイ 118

(74)代理人 弁理士 三俣 弘文

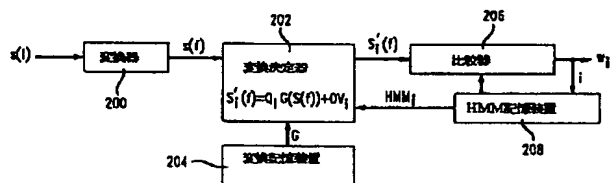
最終頁に続く

(54)【発明の名称】 信号識別方法

(57)【要約】

【目的】 不一致条件下で音声信号認識の平均誤り率を大幅に縮小し、悪環境での使用に適した音声信号認識を実現する。

【構成】 信号（例えば音声信号）と記憶された表現（例えばキーワードの記憶表現）の集合をとり、相対的に、記憶表現をより良好にエミュレートする信号を生じる少なくとも1回の変換を実行する。例えば、これは、3つの技術のうちの1つを使用することによって実現される。第1に、信号が記憶表現のうちの1つによってより良く近似される（例えば接近する）ように信号を変換することがある。第2に、記憶表現のうちの1つが信号をより良く近似するように記憶表現の集合を変換することがある。第3に、信号および記憶表現の集合の両方を変換することがある。記憶表現として、HMM、ニューラルネットワーク、ベクトル量子化表現を用いることができる。



## 【特許請求の範囲】

【請求項 1】 (a) 信号を、記憶表現の集合と比較して、第 1 の類似度値の集合を生成するステップと、

(b) 前記信号の関数を計算するステップと、

(c) 前記信号と、前記関数の少なくとも一部と、前記第 1 の類似度値の集合内の少なくとも 1 つの類似度値に対応する少なくとも 1 つの記憶表現とに基づいて変換を決定するステップと、

(d) 前記変換によって前記信号を変換して変換済み信号を生成するステップと、

(e) 前記変換済み信号を前記記憶表現の集合と比較して第 2 の類似度値の集合を生成するステップと、

(f) 前記第 2 の類似度値の集合に基づいて、前記信号を、特定の記憶表現に類似するものとして識別するステップとからなることを特徴とする信号識別方法。

【請求項 2】 前記記憶表現の集合内の各記憶表現は対応する動作を有し、前記特定の記憶表現に対応する動作を実行するステップをさらに有することを特徴とする請求項 1 の方法。

【請求項 3】 前記信号が音声信号からなることを特徴とする請求項 1 の方法。

【請求項 4】 (a) 前記信号と、前記関数の少なくとも一部と、前記特定の記憶表現とに基づいて新たな変換を決定するステップと、

(b) 前記変換によって前記信号を変換して新たな変換済み信号を生成するステップと、

(c) 前記新たな変換済み信号を前記記憶表現の集合と比較して新たな類似度値の集合を生成するステップと、

(d) 前記新たな類似度値の集合内の少なくとも 1 つの値がしきい値より小さくなるまでステップ a ないし c を反復するステップと、

(e) 前記信号を、前記新たな類似度値の集合内の少なくとも 1 つの値に対応する新たな特定の記憶表現に類似するものとして識別するステップとをさらに有することを特徴とする請求項 1 の方法。

【請求項 5】 前記信号が周波数領域の信号からなることを特徴とする請求項 1 の方法。

【請求項 6】 前記記憶表現の集合が隠れマルコフモデルの集合からなり、前記第 1 の類似度値の集合および前記第 2 の類似度値の集合が尤度値からなることを特徴とする請求項 1 の方法。

【請求項 7】 前記記憶表現の集合がニューラルネットワークの集合からなり、前記第 1 の類似度値の集合および前記第 2 の類似度値の集合がニューラルネットワーク出力値からなることを特徴とする請求項 1 の方法。

【請求項 8】 前記記憶表現の集合がベクトル量子化表現の集合からなり、前記第 1 の類似度値の集合および前記第 2 の類似度値の集合が歪み値からなることを特徴とする請求項 1 の方法。

【請求項 9】 (a) 信号を、記憶表現の集合と比較し

て、第 1 の類似度値の集合を生成するステップと、

(b) 前記信号と、前記第 1 の類似度値の集合内の少なくとも 1 つの類似度値に対応する少なくとも 1 つの記憶表現とに基づいて変換を決定するステップと、

(c) 前記変換によって前記記憶表現の集合を変換して変換済み表現の集合を生成するステップと、

(d) 前記信号を前記変換済み表現の集合と比較して第 2 の類似度値の集合を生成するステップと、

(e) 前記第 2 の類似度値の集合に基づいて、前記信号を、特定の記憶表現に類似するものとして識別するステップとからなることを特徴とする信号識別方法。

【請求項 10】 前記記憶表現の集合内の各記憶表現は対応する動作を有し、前記特定の記憶表現に対応する動作を実行するステップをさらに有することを特徴とする請求項 9 の方法。

【請求項 11】 前記信号が音声信号からなることを特徴とする請求項 9 の方法。

【請求項 12】 (a) 前記信号と、前記特定の記憶表現とに基づいて新たな変換を決定するステップと、

(b) 前記変換によって前記記憶表現の集合を変換して新たな記憶表現の集合を生成するステップと、

(c) 前記新たな記憶表現の集合を前記信号と比較して新たな類似度値の集合を生成するステップと、

(d) 前記新たな類似度値の集合内の少なくとも 1 つの値がしきい値より小さくなるまでステップ a ないし c を反復するステップと、

(e) 前記信号を、前記新たな類似度値の集合内の少なくとも 1 つの値に対応する新たな特定の記憶表現に類似するものとして識別するステップとをさらに有することを特徴とする請求項 9 の方法。

【請求項 13】 前記信号が周波数領域の信号からなることを特徴とする請求項 9 の方法。

【請求項 14】 前記記憶表現の集合が隠れマルコフモデルの集合からなり、前記第 1 の類似度値の集合および前記第 2 の類似度値の集合が尤度値からなることを特徴とする請求項 9 の方法。

【請求項 15】 前記記憶表現の集合がニューラルネットワークの集合からなり、前記第 1 の類似度値の集合および前記第 2 の類似度値の集合がニューラルネットワーク出力値からなることを特徴とする請求項 9 の方法。

【請求項 16】 前記記憶表現の集合がベクトル量子化表現の集合からなり、前記第 1 の類似度値の集合および前記第 2 の類似度値の集合が歪み値からなることを特徴とする請求項 9 の方法。

【請求項 17】 (a) 信号と記憶表現の集合とを繰り返して変換して、信号を、現在の表現の集合内の少なくとも 1 つの記憶表現に近づけるステップと、

(b) 前記信号を、特定の記憶表現に類似するものとして識別するステップとからなることを特徴とする信号識別方法。

【請求項18】 前記記憶表現の集合内の各記憶表現は対応する動作を有し、前記特定の記憶表現に対応する動作を実行するステップをさらに有することを特徴とする請求項17の方法。

【請求項19】 前記信号が音声信号からなることを特徴とする請求項17の方法。

【請求項20】 (a) 前記少なくとも1つの記憶表現に対応する類似度値がしきい値より小さくなるまでステップaを実行することを特徴とする請求項17の方法。

【請求項21】 前記信号が周波数領域の信号からなることを特徴とする請求項17の方法。

【請求項22】 前記記憶表現の集合が隠れマルコフモデルの集合からなり、前記第1の類似度値の集合および前記第2の類似度値の集合が尤度値からなることを特徴とする請求項17の方法。

【請求項23】 前記記憶表現の集合がニューラルネットワークの集合からなり、前記第1の類似度値の集合および前記第2の類似度値の集合がニューラルネットワーク出力値からなることを特徴とする請求項17の方法。

【請求項24】 前記記憶表現の集合がベクトル量子化表現の集合からなり、前記第1の類似度値の集合および前記第2の類似度値の集合が歪み値からなることを特徴とする請求項17の方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、信号認識に関し、特に、悪環境での使用に適した自動信号認識システムの性能を改善することに関する。

【0002】

【従来の技術】信号認識システムの1つのタイプは音声認識システムである。

【0003】音声認識システム（以下単に「システム」という。）は、話者から受け取った入力（すなわち音声）において単語の集合を認識することができる。認識した入力に基づいて、機能が実行される。システムを電話網で使用する場合、その機能は例えば話者をオペレータに接続することである。

【0004】システムは、単語の集合内の各単語を認識するように事前にトレーニングされる。各単語をキーワードという。トレーニングは、トレーニング音声を入力し、キーワードのモデルを形成して記憶することによって行うことができる。

【0005】トレーニングされると、動作時には、システムは入力音声内に含まれるキーワードを認識することができる。システムは、入力音声を記憶されたキーワードモデルと比較することによってこれを行う。入力音声内のキーワードが認識されると、システムはその単語に関係づけられた機能を実行する。

【0006】既知のシステムはかなり高いレベルの精度で入力音声内のキーワードを識別するが、多くの改良の

余地がある。システムの精度は、システムの「平均単語誤り率」によって測定することができる。「平均単語誤り率」は、システムがキーワードを含む発声内で誤ったキーワードを認識するかまたはキーワードが発声されていないときにキーワードを識別する頻度の測度である。

【0007】システムの精度を低下させる1つの要因は「不一致」である。不一致は、システムがある入力システム（例えば、マイクロホンおよびケーブル）を使用してトレーニングされ、別の入力システム（例えば、電話ハンドセットおよび接続された電話網システム）で使用されるときに生じることがある。不一致は、この例の場合、人の声がマイクロホンを通過するときに電話システムの場合と異なる特性を示すために生じるといわれている。

【0008】精度の問題、特に不一致の問題は、これまで注意を受けている。この不一致問題を解決しようとした少なくとも3つの方法がある。

【0009】第1に、話者が入力音声を供給する方法に依存してアプリケーションごとにシステムを訓練することによって、不一致問題を解決することが提案されている。この提案は2つの問題点を有する。第1に、アプリケーションごとにシステムをトレーニングすることは時間を消費し高価である。第2に、システムは、異なる音声入力媒体（例えば、電話およびセルラ電話）がシステムで使用可能である場合には、やはり不一致問題の影響を示すことになる。

【0010】第2に、システムで使用する「プール」モデルを作成することによって不一致問題を解決することが提案されている。複数の入力媒体の効果の混合物を反映するモデルがある。この提案にも2つの問題点がある。第1に、プールモデルは、他のモデルよりも作成するのが高価である（そして一般に時間もかかる）。第2に、与えられた任意の入力システムに対して、その入力システムに基づいて作成されるモデルはプールモデルよりも高い精度を有する。

【0011】第3に、システムに入力される音声信号にオフセット因子を付加することによって不一致問題を解決することが提案されている。付加されるオフセット因子は、信号が受ける歪みを相殺するのに必要なものの推定値を表す。このオフセット因子はコードブック探索によって決定される。コードブックは多くのエン트리（代表的には256個）からなる。これらのエント리는、入力システムの集合からのオフセット値を表す。例えば、コードブックは、第1のタイプの入力システム（例えば、第1のタイプのマイクロホンおよび第1のタイプのチャンネル）に基づく64個のエン트리と、第2のタイプの入力システムに基づく第2の64個のエン트리と、第3のタイプの入力システムに基づく第3の64個のエン트리と、第4のタイプの入力システムに基づく第4の64個のエン트리とを有する。しかし、この提案にも問題

点がある。音声信号として使用される入力第1、第2、第3、または第4のタイプの入力システムではない場合には、使用される入力第256個のコードブックエントリの集合によって正しく特徴づけられないことがある。これによって、音声認識システムの精度は悪くなる。

#### 【0012】

【発明が解決しようとする課題】上記の方法の使用にもかかわらず、不一致条件下での認識に対する平均単語誤り率は多くのアプリケーションでは未だに不満足なものである。不一致問題の解決が必要とされる。

#### 【0013】

【課題を解決するための手段】本発明は、不一致条件下で信号の平均誤り率を大幅に縮小する方法を提供する。本発明の方法は、信号（例えば音声信号）と記憶された表現（例えばキーワードの記憶表現）の集合をとり、相対的に、記憶表現をより良好にエミュレートする信号を生じる少なくとも1回の変換を実行する。例えば、これは、3つの技術のうちの1つを使用することによって実現される。第1に、信号が記憶表現のうちの1つによってより良く近似される（例えば接近する）ように信号を変換することがある。第2に、記憶表現のうちの1つが信号をより良く近似するように記憶表現の集合を変換することがある。第3に、信号および記憶表現の集合の両方を変換することがある。

【0014】本発明に従って形成したシステムは、個々のアプリケーションごとにトレーニングする必要はないという効果がある。

【0015】また、本発明に従って形成したシステムは、プールモデルやコードブックを使用する既知のシステムよりも高い精度を有するという効果がある。

#### 【0016】

【実施例】ここでは、音声認識システムで本発明を実施するという状況で説明する。しかし、当業者には明らかなように、本発明は、物理システムからの物理信号が認識のために記憶表現の集合と比較されるような任意のタイプのシステム（例えば、ファクシミリシステム、衛星システム、光学式文字認識システムなど）で使用可能である。さらに、「強固な音声認識のための確率論的マッチングへの最尤アプローチ(A Maximum Likelihood Approach to Stochastic Matching for Robust Speech Recognition)」という題名の論文を付録Aとして記載する。付録Aは本発明を例示することのみを意図したものである。

【0017】説明を明確にするため、本発明の実施例は個別の機能ブロックからなるものとして表す。これらのブロックが表現する機能は、共用または専用のハードウェアの使用によって実現される。ハードウェアには、ソフトウェアを実行可能なハードウェアも含まれるが、それに限定されるものではない。例えば、図2のブロック

200、202、および206の機能は単一の共用プロセッサによって実現することができる。「プロセッサ」という用語の使用は、ソフトウェアを実行可能なハードウェアのみを指すと解釈してはならない。

【0018】実施例は、AT&TのDSP16またはDSP32Cのようなデジタル信号プロセッサ(DSP)と、以下で説明する動作を実行するソフトウェアを記憶する読み出し専用メモリ(ROM)と、DSPの結果を記憶するランダムアクセスメモリ(RAM)とからなることが可能である。超大規模集積(VLSI)ハードウェア実施例や、カスタムVLSI回路と汎用DSP回路の組合せも実現可能である。

【0019】音声認識システムの状況で本発明を説明する前に、既知の音声認識システムについて簡単に説明する。

【0020】図1を参照すると、音声信号 $S(t)$ が変換器100に入力される。簡単のため、 $S(t)$ は、1個のキーワードを含みそれ以外は含まない発声であると仮定する。変換器100は音声信号 $S(t)$ を周波数領域に変換し、信号 $S(f)$ を生成する。 $S(f)$ はオフセット計算器102に入力される。オフセット計算器102は $Si'(f)$ を出力する。 $Si'(f)$ は $S(f)$ にコードブック104に記憶されているオフセット値 $OVi$ を加えたものに等しい。 $Si'(f)$ は比較器106に入力され、そこで、HMM記憶装置108に記憶されている隠れマルコフモデル(HMM)の集合と比較される。HMM記憶装置108内の各HMMはキーワードに対するHMMである。 $Si'(f)$ を各HMMと比較した後、比較器106は、最も良く $Si'(f)$ に一致するHMMによって表現される単語を出力する。この単語を $Wi$ で表す。次に、このプロセスはコードブック104内の各オフセット値に対して反復される。各 $OVi$ に対して、新たな $Wi$ が比較器106によって決定される。 $Wi$ が各オフセット値 $OVi$ に対して決定されると、最も高い尤度スコアの $Wi$ が、発声に含まれるキーワードを表すものとして選択される。

【0021】既知のシステムについて説明したので、次に、本発明を、3つの異なる実施例における音声認識システムについて説明する。第1実施例は、音声信号の表現（例えばケプストラム表現）を、記憶されている少なくとも1つのHMMにより良く類似するように変換する。第2実施例は、記憶HMMを、音声信号の表現により良く類似するように変換する。第3実施例は、信号および記憶表現の集合の両方を変換し、信号が記憶HMMに類似しているよりも、変換した信号が変換した少なくとも1つのHMMに良く類似するようにする。

【0022】図2を参照して第1実施例について説明する。音声信号 $S(t)$ は変換器200に入力される。音声認識環境では、変換器200は任意のタイプの特徴抽出装置（例えば、スペクトラム分析装置、ケプストラム

分析装置など)である。再び簡単のため、 $S(t)$ は1個のキーワードを含む発声であると仮定する。変換器200は音声信号 $S(t)$ を周波数領域に変換し信号 $S(f)$ を生成する。 $S(f)$ は変換プロセッサ202に10 入力される。これは、図2および図3に示されているように変換決定器とも呼ぶ。変換プロセッサ202は $Si'(f)$ を出力する。 $Si'(f)$ は $QiG(S(f))$ にオフセット値 $OVi$ を加えたものに等しい。 $Qi$ はスケールファクタであり、 $G$ は、変換記憶装置204に記憶されている $S(f)$ の関数である。 $S(f)$ の「関数」は、変換プロセッサで計算される。この関数は必ずしも $Qi$ を含むとは限らない。しかし、この「関数」は、少なくとも3つのこととなる形式を含むほどに広いものである。特に、この「関数」は $QiG(S(f))$ 、 $QiS(f)$ 、または $G(S(f))$ の形式が可能である。 $G$ は任意の線形または非線形の関数であり、その形式はアプリケーションに依存し、信号タイプ、予想される歪み、および特徴抽出方法(例えば、ケプストラム、スペクトラムなど)を含む因子に依存する。 $G$ が上記の因子に基づいて変化するかまたは当業者には容易に明らかとなる。例えば、信号タイプが音声であり、予想される歪みが電話網のチャネル歪みであり、特徴抽出方法がケプストラム分析である場合、 $G(x) = x$ である。

【0023】再び図2を参照すると、変数 $i$ を識別するために、システムの初期通過(初期識別)が必要である。初期通過中には $Si'(f)$ は $S(f)$ に等しく(すなわち、 $Qi=1$ 、 $G(x)=x$ 、および $OVi=0$ )、比較器206は単に $S(f)$ と各HMMの間の比較を実行して第1の尤度値の集合を決定する。この第1尤度値集合に基づいて、 $Wi$ を決定する。 $Wi$ は、初期通過での $Si'(f)$ に最も近いHMM(すなわちHMM<sub>i</sub>)によって表現される単語である。

【0024】さらに図2を参照すると、初期識別後、変換プロセッサ202はHMM記憶装置208からHMM<sub>i</sub>を受信する。HMM記憶装置208は、初期識別後に比較器206からインデックス $i$ を送信されるため、HMM<sub>i</sub>を選択する。変換プロセッサ202は $S(f)$ 、HMM<sub>i</sub>、および変換記憶装置からの入力 $G$ を使用して、関数を計算し、スケールファクタ $Qi$ およびオフセット値 $OVi$ を決定する。 $Qi$ および $OVi$ は付録Aの式32および式33に従って決定される。 $OVi$ 、 $Qi$ および $G$ はこれで既知となるため、変換プロセッサは、 $S(f)$ をとり、上記の式に従ってこれを新たな $Si'(f)$ に変換することが可能となる。

【0025】再び図2を参照すると、 $G$ は数学的に $QiG(S(f))$ と記述される $S(f)$ の関数の一部である。変換(例えば $Si'(f) = Qi(G(S(f)) + OVi)$ )が変換プロセッサ202によって決定されるときに「決定される」必要がある項目は $Qi$ および $OVi$ で

あるため、これは重要である。方程式が2個および未知数が2個あるため、 $Qi$ および $OVi$ はどちらを先に(すなわち $Qi$ が先で $OVi$ が後または $OVi$ が先でその後に $Qi$ )決定することも可能である。この変換を決定することはある状況では関数全体(例えば $Qi$ および $G$ )ではなく関数の一部(例えば $G$ )に基づくこともある。例えば、関数が $QiG(S(f))$ であるとみなされるとする。この関数は計算可能である。しかし、 $OVi$ について先に解き $Qi$ について後に解くことによって変換を決定する場合、関数の少なくとも一部(すなわち $Qi$ )は変換の $OVi$ 部分を決定するのに必要ではない。 $OVi$ が決定されると、 $Qi$ を決定することは自明であり $OVi$ に依存する。従って、変換は関数の少なくとも一部に基づいて決定されることがある。

【0026】再び図2を参照すると、新たな $Si'(f)$ を各HMMと比較し、第2の尤度値の集合を決定する。この第2尤度値集合に基づいて $Wi$ を決定する。この新たな $Wi$ は新たな $Si'(f)$ に最も近いHMMである新たなHMM<sub>i</sub>に対応する。このプロセスは反復されることもあるが、アプリケーションによっては反復する必要はない。例えば、新たな $Wi$ からのインデックス $i$ は次のHMM<sub>i</sub>を識別するHMM記憶装置208に渡される。次のHMM<sub>i</sub>は、 $S(f)$ および $G$ とともに、変換プロセッサに10 入力される。この情報により、変換プロセッサは次の $Qi$ および次の $OVi$ を決定する。

【0027】ある場合には、第2尤度値集合に基づいて、新たな $Si'(f)$ が特定のHMMに最も類似するものとして識別される。しかし、図2について説明したプロセスは反復されることが可能であり、プロセスの追加反復が所望されるような場合もある。このような場合には、プロセスの反復を何回実行するかを決定するいくつかの方法がある。例えば、特定のHMM<sub>i</sub>が現在の $Si'(f)$ に対するあるしきい値より高くなった場合に追加反復を行わないと決定することも可能である。また、ある一定回数の反復より多くを行わないと決定することも可能である。また、以上の2つの方法を組み合わせて、あるしきい値より高いHMM<sub>i</sub>がない限り一定回数の反復を行うことも可能である。このほかにも、当業者には容易に明らかのように、プロセスの反復を何回実行するかを決定するために使用可能な他のいくつかの方法がある。

【0028】図3を参照して第2実施例について説明する。音声信号 $S(t)$ は変換器300に10 入力され、音声信号の周波数表現 $S(f)$ を生成する。再び簡単のため、 $S(t)$ は1個のキーワードを含む発声であると仮定する。 $S(f)$ は比較器302に10 入力される。システムの初期通過中には、 $S(f)$ はHMMの集合(例えばHMM')と比較される。HMM'は、HMM記憶装置304に記憶されているHMMの別の集合に基づいたHMMの集合である。HMM'内の各HMMはHMM記憶

装置のHMMにオフセット値OVを加えたものに等しい。初期通過の場合、OVは0に等しく設定される。初期通過は第1の尤度値の集合を生成する。HMM' 内で最も高い尤度値を与える特定のHMMによって表される単語Wiが識別される。このHMMは、例えば、HMM' 内のi番目のHMMである。こうして、Wiは、音声信号S(t)に最も近い単語の初期決定を表す。

【0029】Wiのインデックス(すなわちi)はHMM記憶装置304に入力される。HMM記憶装置304はこのi番目のHMM(すなわちHMMi)を識別し、それを変換プロセッサ306に入力する。HMMiおよびS(f)に基づいて、変換プロセッサ306はオフセット値OVを決定する。OVは、HMMi、S(f)、および付録Aの式47および式49に基づいて決定される。OVは、HMMi、S(f)、および付録Aの式53および式54に基づいて決定することも可能である。一般に、式47および式49の使用は、式53および式54の使用よりも計算量的にわずかに効率的であるがわずかに有効でない。

【0030】オフセット値をHMM記憶装置の各HMMに加えて新たなHMMの集合(すなわち新たなHMM')を生成する。これは変換プロセッサ306で行われる。こうして、HMM記憶装置304内の記憶HMMの集合が、変換されたHMMの集合(すなわち新たなHMM')を生成するために変換される。

【0031】この新たなHMM'は比較器302でS(f)と比較される。この比較により、第2の尤度値の集合が生成される。この第2尤度値集合に基づいてWiが決定される。この新たなWiは、新たなHMM'内のHMMに対応する。このプロセスは反復されることもあるが、アプリケーションによっては反復する必要はない。

【0032】プロセスの追加反復が所望されることもある。そのような場合、当業者には容易に明らかなように、プロセスの反復を何回実行するかを決定するいくつかの方法がある。

【0033】第3実施例は、第1実施例および第2実施例の両方からの技術を使用する。従って、第3実施例は図2および図3を参照して説明する。

【0034】第3実施例は、信号と、記憶HMMの集合内の少なくとも1つのHMMを互いに接近させる反復法を使用する。例えば、S(t)が図2のシステムに入力されWiが生成される。しかし、インデックスiを図2のHMM記憶装置208に送る代わりに、インデックスiは図3のHMM記憶装置304に送られる。次に、図3のシステムはHMMiおよびS(t)を使用して新たなWiを生成する。第3実施例に従って形成されるシステムによって実行されるプロセスは、1回だけが所望される場合には完了しない可能性もある。しかし、アプリケーションに基づいて追加反復が所望される場合、図3

の新たなWiは新たなインデックスiを図3のHMM記憶装置304ではなく図2のHMM記憶装置208に送る。次に、図2のシステムはプロセスを実行して現在のWiを生成する。次に、現在のWiに対応するiの値が再び図3のHMM記憶装置304に送られる。これは必要なだけ続けることができる。追加反復が所望される場合、当業者には容易に明らかなように、プロセスの反復を何回実行するかを決定するいくつかの方法がある。

【0035】第3実施例を参照すると、2つの事項が当業者には明らかである。第1に、実行する反復回数にかかわらず、第3実施例は、図2または図3に関して説明した技術のいずれを最初に実行することによっても使用可能である。第2に、反復はさまざまな方法で実行可能である。例えば、次のような方法がある。

1. 図2に関して説明した初期識別手順を実行した後、図3に関して説明した手順を実行するか、またはその逆。

2. 図2に関して説明した初期識別手順およびその後の手順を実行した後、図3に関して説明した手順を実行するか、またはその逆。

他の可能性は当業者には明らかである。

【0036】本発明は音声認識システムに関して説明したが、当業者には明らかなように、本発明は、物理信号を記憶された表現の集合と比較するような任意のタイプのシステムで使用可能である。そのようなシステムには、ファクシミリ認識システム、衛星伝送・認識システム、話者認識システム、署名認識システム、および光学式文字認識システムならびにその他の画像または音響認識システムがあるが、これらに限定されるものではない。さらに、各記憶表現は、その記憶表現が信号に類似するものとして識別された場合に実行する対応する動作を有することが可能である。この動作はアプリケーションに依存し、例えば、電話網では発呼者をオペレータに接続することである。当業者には理解されるように、アプリケーションに依存して実行されるさまざまな動作がある。最後に、開示したシステムは単語を表す記憶モデルに関して説明したが、当業者には理解されるように、この記憶モデルは、アプリケーションに依存して、音素、または、音声もしくはデータの他の要素のモデルとすることも可能である。

【0037】[付録A]

#### 要約

試験発声における歪みによって引き起こされる認識性能劣化を縮小するために、試験発声と与えられた音声モデルの集合の間の音響不一致を減少させる最尤(ML)確率論的不一致アプローチを開示する。音声信号が部分語隠れマルコフモデル(HMM)の集合 $\Lambda_x$ によってモデル化されると仮定する。観測される試験発声Yとモデル $\Lambda_x$ の間の不一致は2つの方法、すなわち、(1) Yを、モデル $\Lambda_x$ とより良く一致する発声Xに写像する逆



11

歪み関数  $F_v(\cdot)$  によって、および、(2)  $\Lambda x$  を、発声  $Y$  により良く一致する変換されたモデル (変換済みモデル)  $\Lambda y$  に写像するモデル変換関数  $G_\eta(\cdot)$  によって、縮小することができる。変換  $F_v(\cdot)$  または  $G_\eta(\cdot)$  の関数形を仮定し、期待値最大化 (EM) アルゴリズムを使用して最尤的にパラメータ  $v$  または  $\eta$  を推定する。 $F_v(\cdot)$  または  $G_\eta(\cdot)$  の形の選択は、音響不一致の性質の事前の知識に基づく。

【0038】実験結果を提示して、提案するアルゴリズムの性質を調べ、異なるトランスデューサおよび伝送チャンネルによる不一致の存在するHMMベースの連続音声認識システムの性能の改善におけるこのアプローチの効力を確認する。提案する確率論的マッチングアルゴリズムは急速に収束することがわかる。さらに、不一致条件における認識性能は大幅に改善される一方、一致条件における性能も良好に維持される。不一致条件におけるこの確率論的マッチングアルゴリズムによる平均単語誤り率の縮小は約70%である。

【0039】1. はじめに

最近、悪環境での自動音声認識 (ASR) システムの性能を改善するという問題に関心が集まっている。トレーニング環境と試験環境の間に不一致があると、ASRシステムの性能は劣化する。強固な音声認識の目標は、この不一致の影響を除去して一致条件にできるだけ近い認識性能を実現することである。音声認識では、音声は通常隠れマルコフモデル (HMM)  $\Lambda x$  の集合によってモデル化される。認識中、観測される発声  $Y$  はこれらのモデルを使用して復号される。トレーニング条件と試験条件の間の不一致により、この性能は不一致条件に比較して劣化することが多い。

【0040】トレーニング条件と試験条件の間の不一致は、図4に示した信号空間、特徴空間、またはモデル空間で見ることができる。図4では、 $S$  はトレーニング環境における原音声を示す。トレーニング環境と試験環境の間の不一致は、 $S$  を  $T$  に変換する歪み  $D_1$  によってモデル化される。音声認識では、まず何らかの形の特徴抽出が実行される。その特徴を、トレーニング環境では  $X$  で表し、試験環境では  $Y$  で表す。特徴空間におけるこれら2つの環境の間の不一致は、特徴  $X$  を特徴  $Y$  に変換する関数  $D_2$  によってモデル化される。最後に、特徴  $X$  を使用してモデル  $\Lambda x$  を構築する。モデル空間では、トレーニング環境と試験環境の間の不一致は、 $\Lambda x$  を  $\Lambda y$  に写像する変換  $D_3$  とみなすことができる。不一致の原因には、加法的ノイズ、スペクトラム傾斜およびスペクトラム形成に寄与するチャンネルとトランスデューサの不一致、話者の不一致、異なるアクセント、強制、および異なる話し方がある。最近の多くの研究は、加法的ノイズおよびチャンネルの効果の問題に集中している。

【0041】ノイズに対処するために使用される方法は一般的に3つの大まかなカテゴリーに分けられる。第1

12

のカテゴリーでは、強固な信号処理を使用して、可能な歪みに対する特徴の感度を減少させるものである。1つのアプローチでは、リフタリングのようなスペクトラム形成を行う。その考え方は、低次および高次のケプストラム成分をデエンファサイズすることである。これらの成分は、チャンネルノイズおよび加法的ノイズの効果に敏感であることがわかっているからである。発声からの長時間ケプストラム平均を減算することに基づく方法も提案されている。この考え方は、チャンネルによる不一致を除去するために一般的に使用される。またこの第1のカテゴリーには、スペクトラムシーケンスをハイパスフィルタリングして緩変動チャンネルの効果除去する方法もある。聴覚モデリングに基づく方法では、信号処理を使用して人間の耳の処理を模倣し、より強固な特徴が得られることを期待する。音声特徴へのノイズの影響を縮小することができる信号制限プリプロセッサの使用も知られている。ノイズの影響を縮小するもう1つの方法は、トレーニングデータに環境ノイズの一部を注入してシステムを再トレーニングすることである。この技術はディザリングに似ている。また、スペクトラム減算に基づく方法もある。この方法では、ノイズパワースペクトラムの推定値を各音声フレームから減算する。この第1のカテゴリーのアプローチは、一般に何らかの形の強固な特徴前処理を含むので、特徴空間 (図4) で作用するとみなすことができる。

【0042】第2のカテゴリーは、何らかの最適基準を使用して、明瞭な音声の関数の推定値を形成することである。音声スペクトラムの関数の最小平均2乗誤差 (MMSE) 推定に基づく定式化において、破壊的ノイズが独立のガウス過程であると仮定するものが知られている。さらに、各スペクトラムビンは別個に推定される。個々のビンは独立であると仮定したためである。音声分布をガウシアン混合 (ミクスチャ) としてモデル化して、別々のスペクトラムビン間の相関を各ミクスチャの対角共分散行列でモデル化する。最後に、音声の時間構造は、隠れマルコフモデル (HMM) によって音声をモデル化することにより考慮される。これらのアプローチは、信号空間と特徴空間のいずれの表現を推定しているかに依存して、音声強化として信号空間で、または、スペクトラム補償として特徴空間で、見ることができる。

【0043】第3のカテゴリーでは、ノイズをモデル化し直接認識プロセスに組み込む。このアプローチの1つは、ノイズマスキングに基づく。この考え方では、信号エネルギーがある適当なノイズレベル以下になった場合に、そのノイズレベルでフィルタバンクエネルギーを置き換える。こうして、ノイズによって顕著に破壊された情報は無視される。モデル分解と呼ばれるもう1つのアプローチでは、音声およびノイズの別々のHMMをトレーニングデータからトレーニングする。認識中には、これらの2つのモデルの結合状態空間でビタビ検索を実行

13

する。この方法はかなり良好に動作することが示されているが、音声およびノイズの両方に対する精度の良いモデルが必要である。このモデル分解アプローチは上記のアプローチに似ている。しかし、原音声のパラメータは、認識中のノイズのある音声から推定される。信号とノイズのモデルの間のさらに一般的な相互作用を許容した、ノイズのあるデータから原音声パラメータを推定する問題は良く研究されている。この場合、信号はガウシアンミクスチャとしてモデル化されると仮定される。このカテゴリーのさらにもう1つの方法では、HMMパラメータを推定する前に、トレーニング音声のエネルギー等高線の最尤(ML)推定を使用して音声を正規化する。試験中には、明瞭な利得パラメータのML推定がノイズのある音声から計算され、それを使用して、音声モデルのパラメータを正規化する。音声認識に対するミニマックス法は既知であり、認識器は、トレーニング中に推定された値の近傍をHMMパラメータが占有するようにすることによってさらに強固となる。これらのアプローチは、図4に示したモデルにおける可能な歪みを扱うモデル空間で作用するとみなされる。

【0044】話者とチャンネルの適応に関する最近の研究では、各話者を基準話者に変換する固定バイアスを推定した後、推定されたバイアスを各音声フレームから減算する。類似のアプローチが、音声ベクトル量子化(VQ)コードブックによってモデル化されるような音声認識においてチャンネル不一致を推定するために使用されている。チャンネル不一致を推定するもう1つのアプローチとして、推定が、2つのチャンネルの平均対数スペクトラム間の差に基づいて行うことが提案されている。

【0045】ここに開示するのは、強固な音声認識のための確率論的マッチングへのMLアプローチである。本方法では、発声の認識中に、MLアプローチを使用することによって、観測発声と原音声モデルの間の不一致を縮小する。この不一致は少なくとも2つの方法で縮小することができる。第1に、歪んだ特徴 $Y$ を原特徴 $X$ の推定値に写像し、原モデル $\Lambda x$ を認識に用いることができるようにすることが可能である。第2に、原モデル $\Lambda x$ を、観測発声 $Y$ により良く一致する変換済みモデル $\Lambda y$ に写像することが可能である。第1の写像は特徴空間で作用し、第2の写像はモデル空間で作用する(図4)。これらの写像を、特徴空間では $F_v(Y)$ で表し、モデル空間では $G_\eta(\Lambda x)$ で表す。ただし $v$ および $\eta$ は推定すべきパラメータである。これらの写像の関数形は、音響不一致の性質に関する事前の情報に依存する。次に、与えられたモデル $\Lambda x$ に対して観測音声 $Y$ の尤度を最大化して歪みによる不一致を減少させるようにこれらの関数のパラメータ $v$ または $\eta$ を推定する。目標は認識を改善するために不一致を縮小することであり、 $\Lambda x$ は認識に使用するモデルであるので、パラメータ $v$ および $\eta$ を推定するための音声モデルとしてHMMの $\Lambda x$ を使

14

用することは直感的に興味のあることである。MLパラメータ推定は、反復して尤度を改善する期待値最大化(EM)アルゴリズムを使用して解かれる。本発明の確率論的マッチングアルゴリズムは与えられた試験発声および与えられた音声モデルの集合のみに作用するため、実際の試験前の不一致の推定にトレーニングは不要である。上記の2つの方法をともに使用して、不一致の効果を縮小する第3の方法とすることも可能である。

【0046】異なるトランスデューサおよびチャンネルによる不一致の存在下で連続音声認識システムの性能を改善するアプローチの効力を示すために実験結果を提示する。この不一致は、固定バイアス(特徴空間において)として、およびランダムバイアス(モデル空間において)としての両方でモデル化される。提案するアプローチは、不一致条件で単語誤り率を約70%縮小し、一致条件下での性能を維持した。本発明のアルゴリズムは、急速に収束する(反復2回以内)こともわかった。

【0047】以下の説明の構成は次の通りである。第2節で、変換 $F_v(\cdot)$ および $G_\eta(\cdot)$ のパラメータの最尤推定に対する一般的な枠組みを説明する。第3節で、特徴空間変換の場合を説明する。特に、未知パラメータに関して線形であるが、観測値に関して非線形な逆歪み関数のパラメータの推定値に対する表式を導出する。特別な場合として、加法的バイアスモデルを考察する。第4節で、変換をモデル空間で見る。特に、ランダム加法的バイアス歪みの場合を考察する。

#### 【0048】2. 確率論的マッチングの構成

パターン認識では、トレーニングモデルの集合 $\Lambda x = \{\lambda_{xi}\}$ (ただし、 $\lambda_{xi}$ は $i$ 番目のクラスのモデルである。)、および試験データの集合 $Y = \{y_1, y_2, \dots, y_I\}$ が与えられた場合に、 $Y$ に埋め込まれた事象の列 $W = \{W_1, W_2, \dots, W_L\}$ を認識することが所望される。連続音声認識の場合、例えば、 $\lambda_{xi}$ は $i$ 番目の部分語HMM単位に対応し、 $Y$ は特定の試験発声に対応する。その場合 $W$ は復号された単音または単語の列となる。モデル $\Lambda x$ をトレーニングする際には、トレーニングデータの集合に制限される。残念ながら、このトレーニングデータと試験データ $Y$ の間に不一致が存在することがあり、このことは、認識される列 $W$ における誤りを引き起こす。この不一致は、もとの信号空間、特徴空間、またはモデル空間(図4)で見ることができる。図中、関数 $D(\cdot)$ は原空間に対応する歪みのある空間に写像する。不一致の原因には、信号中の歪み、信号の不完全な特徴づけ、不十分な量のトレーニングデータ、または不適切なモデル化および推定誤差がある。以下では、トレーニング音声データと試験音声データの不一致による音声認識性能の劣化の問題を考える。この不一致は、マイクロホンとチャンネルの不一致、トレーニングと試験の環境の相違、話者および話し方またはアクセントの相違、またはこれらの任意の組合せによる可能性

がある。

【0049】音声認識では、モデル $\Lambda_X$ を使用して、最 \*

\*大事後 (MAP) デコーダ

【数1】

$$W' = \arg \max_W p(Y|W, \Lambda_X) P(W) \quad (1)$$

を用いてYを復号する。ただし、第1項は単語列Wが与えられた場合にYを観測する尤度であり、第2項は単語列Wの事前確率である。この第2項は、許容単語列の集合に制約を加える言語モデルと呼ばれる。トレーニング環境と試験環境の間の不一致により、式1によって評価される $\Lambda_X$ が与えられたときのYの尤度に対応する不一致があり、復号された列W'に誤りを引き起こす。この※

※不一致を減少させることにより認識性能が改善される。

【0050】特徴空間において、歪み関数が原発声 $X = \{x_1, x_2, \dots, x_T\}$ を観測値の列 $Y = \{y_1, y_2, \dots, y_T\}$ に写像するとする。この歪みが可逆である場合、次のような逆関数 $F_v$ でYを原発声Xに写像することができる。

【数2】

$$X = F_v(Y) \quad (2)$$

ただし、 $v$ は逆歪み関数のパラメータである。あるいは、モデル空間において、パラメータ $\eta$ を有し $\Lambda_X$ を変 \*

★換済みモデル $\Lambda_Y$ に写像する変換 $G_\eta$ を考える。

【数3】

$$\Lambda_Y = G_\eta(\Lambda_X) \quad (3)$$

Yと $\Lambda_X$ の間の不一致を減少させる1つのアプローチは、モデル $\Lambda_X$ が与えられたときの式1のYとWの結合尤度を最大にするパラメータ $v$ または $\eta$ および単語列W ☆ 20

☆を見つけることである。すなわち、特徴空間では、次のような $v'$ を見つけることが必要である。

【数4】

$$(v', W') = \arg \max_{(v, W)} p(Y, W|v, \Lambda_X) \quad (4)$$

これは次式と等価である。

【数5】

$$(v', W') = \arg \max_{(v, W)} p(Y|W, v, \Lambda_X) P(W) \quad (5)$$

対応して、モデル空間では、次のような $\eta'$ を見つけることが必要である。

◆【数6】

$$(\eta, W') = \arg \max_{(\eta, W)} p(Y, W|\eta, \Lambda_X) \quad (6)$$

これは次式と等価である。

【数7】

$$(\eta', W') = \arg \max_{(\eta, W)} p(Y|W, \eta, \Lambda_X) P(W) \quad (7)$$

式5における変数 $v$ とWに関する、または、式7における $\eta$ とWに関するこの同時最大化は、 $v$ または $\eta$ を固定してWについて最大化し、その後、Wを固定して $v$ または $\eta$ について最大化することを反復することにより行われる。この手続きを、概念的に、特徴空間について図5に、また、モデル空間について図6に示す。 \*

\*【0051】Wを見つけるプロセスは多くの研究者によって扱われている。パラメータ $v$ および $\eta$ を見つける問題は興味がある。表式を簡単にするため、W依存性を除去し、式5および7に対応する最尤推定問題を

【数8】

$$v' = \arg \max_v p(Y|v, \Lambda_X) \quad (8)$$

および

【数9】

$$\eta' = \arg \max_\eta p(Y|\eta, \Lambda_X) \quad (9)$$

のように書く。

分語HMMの集合であると仮定する。 $i, j =$

【0052】これを調べるため、 $\Lambda_X$ が左右連続密度部

50  $1, \dots, N$ に対して、状態 $i$ から $j$ への遷移確率を $a$

17

$i, j$  で表し、状態  $i$  に対する観測値密度  $p_x(x|i)$  は \* であると仮定する。  
次式によって与えられるようなガウシアン・のミクスチャ \* 【数10】

$$p_x(x|i) = \sum_{j=1}^M w_{i,j} N(x; \mu_{i,j}, C_{i,j}) \quad (10)$$

ただし  $M$  はミクスチャの数であり、 $w_{i,j}$  は状態  $i$  におけるミクスチャ  $j$  の確率であり、 $N$  は次式によって与え ※ られる正規分布である。 【数11】

$$N(x; \mu_{i,j}, C_{i,j}) = \frac{1}{(2\pi)^{D/2} |C_{i,j}|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_{i,j})^T C_{i,j}^{-1} (x - \mu_{i,j})\right) \quad (11)$$

ただし  $D$  は特徴ベクトル  $x$  の次元であり、 $C_{i,j}$  および  $\mu_{i,j}$  は状態  $i$  におけるミクスチャ  $j$  に対応する共分散行列および平均値ベクトルである。

【0053】  $S = \{s_1, s_2, \dots, s_T\}$  を、モデルの集合  $\Lambda_X$  に対するすべての可能な状態列の集合とし、 $C \star$

$\star = \{c_1, c_2, \dots, c_T\}$  を、すべてのミクスチャ成分列の集合とする。すると、式8は次のように書くことができる。

【数12】

$$v' = \arg \max_v p(Y|v, \Lambda_X) = \arg \max_v \sum_S \sum_C p(Y, S, C|v, \Lambda_X) \quad (12)$$

同様に、式9は次のように書くことができる。

20 【数13】

$$\eta' = \arg \max_{\eta} \sum_S \sum_C p(Y, S, C|\eta, \Lambda_X) \quad (13)$$

【0054】 一般に、 $v'$  または  $\eta'$  を直接推定することは容易ではない。しかし、ある  $F_v$  および  $G_\eta$  に対しては、EMアルゴリズムを使用して現在の推定値を反復的に改善し、式12および13中の尤度が反復ごとに増大するように新たな推定値を得ることができる。次の2つの節で、特徴空間変換  $F_v$  のパラメータ  $v$ 、およびモデル空間変換  $G_\eta$  のパラメータ  $\eta$  の推定値を見つけるためのEMアルゴリズムの応用をそれぞれ説明する。 ☆

### ☆【0055】 3. 特徴空間変換 $F_v$ の推定

この節では、EMアルゴリズムを使用して式8の推定値  $v'$  を見つける。EMアルゴリズムは2ステップ反復手続きである。第1ステップは、期待値ステップ (Eステップ) と呼ぶ。この第1ステップで、次式によって与えられる補助関数を計算する。

30 【数14】

$$Q(v'|v) = E\{\log p(Y, S, C|v', \Lambda_X) | Y, v, \Lambda_X\} \\ = \sum_S \sum_C p(Y, S, C|v, \Lambda_X) \log p(Y, S, C|v', \Lambda_X) \quad (14)$$

第2ステップは、最大化ステップ (Mステップ) と呼ぶ。この第2ステップで、 $Q(v'|v)$  を最大にする ◆

◆  $v'$  の値を見つめる。すなわち、

【数15】

$$v' = \arg \max_v Q(v'|v) \quad (15)$$

である。 $Q(v'|v) \geq Q(v|v)$  である場合、 $p(Y|v', \Lambda_X) \geq p(Y|v, \Lambda_X)$  となることを示すことができる。従って、式14および15のEステップおよびMステップを反復して適用した場合に尤度が非減少であることが保証される。反復は、尤度の増大がある所定のしきい値未満になるまで継続される。 \*

\* 【0056】 一般に、式2の関数  $F_v(\cdot)$  は  $Y$  のブロックを異なるサイズの  $X$  のブロックに写像することができる。しかし、簡単のため、この関数は、次式のように、 $Y$  の各フレームを対応する  $X$  のフレームに写像するようなものであると仮定する。

【数16】

$$x_i = f_v(y_i) \quad (16)$$

上記のような連続密度HMMによって与えられる  $\Lambda_X$  に 50 より、この補助関数は次のように書き換えることができる

る。

【数17】:

$$Q(v'|v) = \sum_S \sum_C p(Y, S, C | v, \Lambda_X) \log \prod_{t=1}^T a_{s_{t-1}, s_t} w_{s_t, c_t} p_y(y_t | s_t, c_t, v', \Lambda_X) \quad (17)$$

ただし、 $p_y(y_t | s_t, c_t, v, \Lambda_X)$  は、ランダム変数  $y_t$  の確率密度関数である。これは、式11によつて与えられるランダム変数  $x_t$  の密度関数と関係  $x_t = f * v(y_t)$  から導出することができる。  $y_t$  の密度は次のように書くことができる。

$$p_y(y_t | s_t, c_t, v, \Lambda_X) = \frac{N(f_v(y_t); \mu_{s_t, c_t}, C_{s_t, c_t})}{|J_v(y_t)|} \quad (18)$$

ただし、 $J_v(y_t)$  は、 $(i, j)$  成分が次式によつて与えられるようなヤコビ行列である。

※【数19】

$$J_{v, j} = \frac{\partial y_{t, j}}{\partial f_{v, j}(y_t)} \quad (19)$$

ただし、 $y_{t, i}$  は  $y_t$  の第  $i$  成分であり、 $f_{v, j}(y_t)$  は  $f_v(y_t)$  の第  $j$  成分である。さらに、式17は次★

★のように書き換えることができる。

【数20】

$$Q(v'|v) = \sum_S \sum_C p(Y, S, C | v, \Lambda_X) \cdot \sum_{t=1}^T \left\{ \log a_{s_{t-1}, s_t} + \log w_{s_t, c_t} + \log \frac{N(f_v(y_t); \mu_{s_t, c_t}, C_{s_t, c_t})}{|J_v(y_t)|} \right\} \quad (20)$$

これはさらに次式のように書ける。

【数21】

$$\begin{aligned} Q(v'|v) = & \sum_{n=1}^N p(Y, s_1 = n | v, \Lambda_X) \log a_{s_0, n} \\ & + \sum_{t=2}^T \sum_{n=1}^N \sum_{l=1}^N p(Y, s_t = n, s_{t-1} = l | v, \Lambda_X) \log a_{l, n} \\ & + \sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M p(Y, s_t = n, c_t = m | v, \Lambda_X) \log w_{n, m} \\ & + \sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M p(Y, s_t = n, c_t = m | v, \Lambda_X) \log N(f_v(y_t); \mu_{n, m}, C_{n, m}) \\ & - \sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M p(Y, s_t = n, c_t = m | v, \Lambda_X) \log |J_v(y_t)| \end{aligned} \quad (21)$$

ここで、 $a_{s_0, n}$  は状態  $n$  の初期確率である。式21 40 ★ように書くことができる。

の補助関数を計算する際には、 $v'$  を含む項にのみ興味

【数22】

がある。従つて、式11を用いて、この補助関数を次の★

$$Q(v'|v) = \sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \gamma_t(n, m) \left[ -\frac{1}{2} (f_v(y_t) - \mu_{n, m})^T C_{n, m}^{-1} (f_v(y_t) - \mu_{n, m}) - \log |J_v(y_t)| \right] \quad (22)$$

ただし、 $\gamma_t(n, m) = p(Y, s_t = n, c_t = m | v, \Lambda_X)$  は、 $Y$  と、観測値  $y_t$  を生成する状態  $n$  からのミクスチャ  $m$  との結合尤度である。次のように、前方後方アルゴリズムを使用して確率  $\gamma_t(n, m)$  を計算す

ることができる。

【数23】

$$\gamma_i(n, m) = \alpha_i(n) \beta_i(n) \frac{w_{n,m} N(f_v(y_i); \mu_{n,m}, C_{n,m})}{\sum_{j=1}^M w_{n,j} N(f_v(y_i); \mu_{n,j}, C_{n,j})} \quad (23)$$

ただし、

【数24】

$$\alpha_i(n) = p(y_1, y_2, \dots, y_i, s_i = n | v, \Lambda_X) \quad (24)$$

【数25】

10

$$\beta_i(n) = p(y_{i+1}, y_{i+2}, \dots, y_T | v, s_i = n, \Lambda_X) \quad (25)$$

である。前方後方アルゴリズムを使用して、 $\alpha_i(n)$  および  $\beta_i(n)$  を反復して計算することができる。

\*は、式22の右边を  $v'$  で微分してその零点を解くことにより、陽に解を導くことができる。すなわち、次のような  $v'$  を見つけることができる。

【0057】  $v'$  に関する  $Q(v' | v)$  の最大値を見

【数26】

つけるために、勾配上昇アルゴリズムのような任意の山

登り法を使用することができる。しかし、場合によって \*

$$\frac{\partial}{\partial v'} \sum_{i=1}^T \sum_{n=1}^N \sum_{m=1}^M \gamma_i(n, m) \quad (26)$$

$$\left[ -\frac{1}{2} (f_v(y_i) - \mu_{m,n})^T C_{n,m}^{-1} (f_v(y_i) - \mu_{m,n}) - \log |J_v(y_i)| \right] = 0$$

【0058】 式8の尤度の最大にするEMアルゴリズムは、式22から  $Q(v' | v)$  を計算 (Eステップ) した後、式26から  $v'$  を見つける (Mステップ) ことによって実行される。その後、この値を式22の  $v$  に代入し、アルゴリズムは反復して実行される。

※で、分節MLアプローチを使用して、式12の尤度  $p(Y | v, \Lambda_X)$  を直接最大にする代わりに、観測値と状態列の結合尤度  $p(Y, S | v, \Lambda_X)$  を最大にすることも可能である。この場合、反復推定手続きは次のようになる。

【0059】 分節k平均アルゴリズムとのアナロジー ※

【数27】

$$S^l = \arg \max_S p(Y, S | v^l, \Lambda_X) \quad (27)$$

【数28】

$$v^{l+1} = \arg \max_v p(Y, S^l | v, \Lambda_X) \quad (28)$$

こうして、まず最尤状態列  $S^l$  を見つけた後、この状態列の条件付きで発声  $Y$  の尤度を最大にする  $v^{l+1}$  を見

★用して  $v^{l+1}$  を見つけることが可能である。容易に示されるように、上記のEM手続きは、 $\gamma_i(n, m)$  が次式によって定義されることを除いてはやはり成り立つ。

ける。ビタビアルゴリズムを使用して、最適な状態列  $S^l$  を見つけることが可能であり、EMアルゴリズムを使 ★40

【数29】

$$\gamma_i(n, m) = \begin{cases} \frac{w_{n,m} N(f_v(y_i); \mu_{n,m}, C_{n,m})}{\sum_{j=1}^M w_{n,j} N(f_v(y_i); \mu_{n,j}, C_{n,j})} & s_i^l = n \text{ の場合} \\ 0 & \text{それ以外の場合} \end{cases} \quad (29)$$

【0060】 簡単のため、 $f_v(y_t)$  は各成分に別個に作用し (すなわち、 $x_{t,i} = f_{v,i}(y_{t,i})$ )、共分散行列  $C_{n,m}$  は対角形である (すなわち、 $C_{n,m} = \text{diag}(\sigma_{n,m}^2)$ ) と仮定する。以下では、表現を容易にす 50

るため、ベクトルの第  $i$  成分を表す添字  $i$  の参照を省略する。次の形の関数を考える。

【数30】

$$f_v(y_t) = ag(y_t) + b \quad (30)$$

ただし、 $g(y_t)$  は  $y_t$  についての既知の（おそらくは非線形の）微分可能関数であり、 $v = \{a, b\}$  は既知パラメータの集合である。すると、式22の補助関数は\*

\*次のように書くことができる。  
【数31】

$$Q(a', b' | a, b) = \sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \gamma_t(n, m) \left[ -\frac{1}{2} \frac{(a'g(y_t) + b' - \mu_{m,n})^2}{\sigma_{n,m}^2} + \log a' \right] \quad (31)$$

式31を  $a'$  および  $b'$  のそれぞれに関する微分をとつ 10 ※【数32】

て0とおくことにより、

$$\sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \gamma_t(n, m) \left[ \frac{1}{a'} - \frac{(a'g(y_t) + b' - \mu_{m,n})g(y_t)}{\sigma_{n,m}^2} \right] = 0 \quad (32)$$

および

$$\sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \gamma_t(n, m) \frac{a'g(y_t) + b' - \mu_{m,n}}{\sigma_{n,m}^2} = 0 \quad (33)$$

を得る。式32および33は、推定値  $a'$  および  $b'$  に 20 ★0の特別の場合、すなわち、  
【数34】

【0061】次に、加法的バイアス  $b_t$  に対応する式3 ★

$$x_t = y_t - b_t \quad (34)$$

の場合を考える。各成分について  $a=1$ 、 $g(y_t) = y_t$ 、および  $b = -b_t$  の場合、式34は式30と等価となる。観測値がスペクトラム領域にある場合、 $b_t$  は加法的ノイズスペクトラムと解釈することができる。一方、観測値がケプストラム領域すなわち対数エネルギー領域にある場合、 $b_t$  は例えばトランスデューサまたは

☆は、区分的一定バイアス、あるいは信号状態依存バイアスがある。あるいは、バイアスを確率的にモデル化し、歪みをモデル空間（詳細は第4節）で見ること可能である。この節では、状態依存バイアスおよび固定バイアスの場合について考察する。

【0062】バイアス  $b_t$  は、ある発声に対して固定したものとして、または、時間とともに変動するものとして、モデル化することができる。時変バイアスの例に ☆

30 【0063】まず状態依存の場合を考える。バイアスはHMM状態ごとに変動する。各音声状態  $n$  に対応して特定のバイアス項  $b_n$  があると仮定する。式22の補助関数を次のように書くことができる。

$$Q(b' | b) = \sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \gamma_t(n, m) \left[ -\sum_{j=1}^D \frac{(y_{t,j} - b'_{n,j} - \mu_{m,n,j})^2}{2\sigma_{n,m,j}^2} \right] \quad (35)$$

ただし、 $b = \{b_1, \dots, b_n, \dots, b_N\}$  である。式26の再推定手続きには、式35の  $b'_{n,i}$  に関する導関数を計算して0と等置することが必要である。その結 ◆

◆果次式を得る。

40 【数36】

$$\sum_{t=1}^T \sum_{m=1}^M \gamma_t(n, m) \frac{y_{t,j} - b'_{n,j} - \mu_{m,n,j}}{\sigma_{n,m,j}^2} = 0 \quad (36)$$

これは次式を与える。

【数37】

$$b'_{n,j} = \frac{\sum_{t=1}^T \sum_{m=1}^M \gamma_t(n,m) \frac{y_{t,j} - \mu_{m,n,j}}{\sigma_{n,m,j}^2}}{\sum_{t=1}^T \sum_{m=1}^M \frac{\gamma_t(n,m)}{\sigma_{n,m,j}^2}} \quad (37)$$

単一の固定バイアス  $b$  の場合、同様にして次式を示すことができる。 \* 【数38】

$$b'_i = \frac{\sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \gamma_t(n,m) \frac{y_{t,i} - \mu_{m,n,i}}{\sigma_{n,m,i}^2}}{\sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \frac{\gamma_t(n,m)}{\sigma_{n,m,i}^2}} \quad (38)$$

【0064】式37からわかるように、推定すべき状態依存バイアス項が多い場合に、小さいサンプルの効果によって推定の問題が起こることがある。しかし、状況によっては、状態依存バイアスには意味がある。例えば、線形フィルタリングに対する加法的ケプストラムバイアスは信号対ノイズ比 (SNR) が高い場合にのみ妥当である。SNRが低いときには、ノイズが優勢となり、チャネルフィルタリングに対する加法的バイアスモデルは不正確となる。これを扱う1つの方法は、そのバイアスがSNR依存であると仮定し、異なるSNR範囲に応じて異なるバイアスを推定することである。このようなアプローチを実現したもの1つに、音声および背景の分節に異なるバイアスを推定するものがある。これは、不一致の一部が電話チャネルによって引き起こされるよう※

$$y_t = f(x_t, b_t) \quad (39)$$

すると、 $x_t$ と $b_t$ が独立である場合、 $y_t$ の確率密度関数 (pdf) を次のように書くことができる。 ★

$$p(y_t) = \iint_{H_t} p(x_t) p(b_t) dx_t db_t \quad (40)$$

ただし、 $H_t$ は式39によって与えられる曲線である。

【0067】前のように、 $X$ はHMMの集合 $\Lambda_X$ によってモデル化される。 $B$ の統計モデルが $\Lambda_B$ によって与えられるとする。 $\Lambda_B$ はHMMまたは混合ガウス密度とす ☆

$$p(b_t) = N(b_t; \mu_b, \sigma_b^2) \quad (41)$$

さらに、式34のときのように、加法的バイアスという特別の場合を考え、曲線 $H_t$ が次式によって与えられる ◆

$$y_t = x_t + b_t \quad (42)$$

これらの仮定のもとで、 $\Lambda_Y$ の構造は $\Lambda_X$ の構造と同じく保たれる。 $\Lambda_Y$ の各ミクスチャ成分の平均および分散は、次のように、平均 $\mu_b$ および分散 $\sigma_b^2$ を、 $\Lambda_X$ における  $\mu_y = \mu_x + \mu_b$

※な場合に有用であることがわかっている。これはおそらく電話チャネルに存在する加法的ノイズによるものである。その結果の詳細は第5節で説明する。

#### 【0065】4. モデル空間変換 $G_\eta$ の推定

前節では、歪みのある音声 $y$ が原音声 $x$ の定関数であると仮定した。この節では、歪みはランダムであるとみなし、それをモデル空間で見る。すなわち、歪みは原モデル $\Lambda_X$ を歪みのあるモデル $\Lambda_Y$ に変換する (図4)。

【0066】観測値列 $Y = \{y_1, \dots, y_T\}$ が、原発声 $X = \{x_1, \dots, x_T\}$ および歪み列 $B = \{b_1, \dots, b_T\}$ と次式によって関係しているとす

る。 【数39】

★【数40】

☆ることが可能である。この考察では、 $\Lambda_B$ は、次のような、対角形共分散行列を有する単一のガウス密度であると仮定する。

【数41】

◆とする。

【数42】

る対応するミクスチャ成分の平均および分散に加えることによって導出される。

【数43】

(43)



【数44】

$$\sigma_y^2 = \sigma_x^2 + \sigma_b^2 \quad (44)$$

式43および44は、式3のモデル変換 $G_\eta(\cdot)$ を定義し、パラメータ $\eta$ は $\mu_b$ および $\sigma_b^2$ によって与えられる。 $\Lambda_b$ がHMMや混合ガウス密度のようにさらに複雑である場合、 $\Lambda_Y$ の構造は、状態およびミクスチャ成分の数が異なるという点で、 $\Lambda_X$ の構造とは異なることも\*

\*ある。

【0068】ここで、式13のパラメータ推定問題を次のように書くことができる。

【数45】

$$\begin{aligned} \eta' &= (\mu_b', \sigma_b'^2) = \arg \max_{\mu_b, \sigma_b^2} \sum_S \sum_C p(Y, S, C | \eta, \Lambda_X) \\ &= \arg \max_{\mu_b, \sigma_b^2} \sum_S \sum_C \prod_{t=1}^T a_{s_{t-1}, s_t} w_{s_t, c_t} N(y_t; \mu_{s_t, c_t} + \mu_b, \sigma_{s_t, c_t}^2 + \sigma_b^2) \end{aligned} \quad (45)$$

再び、EMアルゴリズムを使用して反復して $\mu_b$ および $\sigma_b^2$ を推定することができる。容易に示されるように、式21に対応する補助関数は次のように書くことができ※

※る。

【数46】

$$Q(\eta' | \eta) = - \sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \gamma_t(n, m) \sum_{i=1}^D \left[ \frac{1}{2} \log(\sigma_{n, m, i}^2 + \sigma_{b_i}^2) + \frac{(y_{t, i} - \mu_{b_i}' - \mu_{m, n, i})^2}{2(\sigma_{n, m, i}^2 + \sigma_{b_i}^2)} \right] \quad (46)$$

この関数を $\eta' = (\mu_b', \sigma_b'^2)$ に関して最大化することによって、 $\sigma_b'^2$ に対する閉じた表式は得られないが、式38と類似の表式が $\mu_b'$ について次のように★

★得られる。

【数47】

$$\mu_{b_i}' = \frac{\sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \gamma_t(n, m) \frac{y_{t, i} - \mu_{m, n, i}}{\sigma_{n, m, i}^2 + \sigma_{b_i}^2}}{\sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \frac{\gamma_t(n, m)}{\sigma_{n, m, i}^2 + \sigma_{b_i}^2}} \quad (47)$$

【0069】 $\sigma_b'^2$ を推定する問題への1つのアプローチは、分散 $\sigma_b^2$ が信号状態依存であり、信号分散と次式☆

☆によって関係していると仮定することである。

【数48】

$$\sigma_{b, n, m, i}^2 = \alpha_i \sigma_{n, m, i}^2 \quad (48)$$

ただし、 $\alpha_i$ は分散の第 $i$ 成分に対するスケールファクタである。式48を式46に代入し、 $\alpha_i$ に関して最大◆40

◆化することによって、次式を得る。

【数49】

$$1 + \alpha_i = \frac{1}{T} \sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \gamma_t(n, m) \frac{(y_{t, i} - \mu_{m, n, i} - \mu_{b_i}')^2}{\sigma_{n, m, i}^2} \quad (49)$$

式42の場合のような物理的意味づけを式48の仮定に対してすることは容易にはできないが、これにより式47および49に示されるようにパラメータ $\mu_b$ および $\sigma_b^2$ の閉じた推定が得られる。 $\alpha_i > -1$ は分散膨張( $\alpha >$

0)および分散収縮( $\alpha < 0$ )の両方に対応する。

【0070】式42と整合する別のアプローチは、尤度 $p(Y | \eta, \Lambda_X)$ を次のように書くことである。

【数50】

$$p(Y|\eta, \Lambda_X) = \sum_S \sum_C \int_{H_T} p(X, B, S, C|\eta, \Lambda_X) dX dB \quad (50)$$

ただし、 $\int_{H_T}$ は次式によって与えられるT重積分である。 \* 【数51】

$$\int_{H_T} dX dB = \prod_{t=1}^T \int_{H_t} dx_t db_t \quad (51)$$

対応して、新たな補助関数を次のように定義することが \* 【数52】  
※10  
できる。

$$Q(\eta'|\eta) = \sum_S \sum_C \int_{H_T} p(X, B, S, C|\eta, \Lambda_X) \log p(X, B, S, C|\eta', \Lambda_X) dX dB \quad (52)$$

このアプローチは、ノイズのある観測値が与えられた場合に原音声モデルのパラメータの推定に対する一般的表式を導出するためにとられている。ただし、音声および歪みは両方とも混合ガウス密度によってモデル化される。この節で考察した問題は、歪みのある音声を与えられた場合に歪みのあるモデルのパラメータを見つける問★

★題の逆である。音声モデル $\Lambda_X$ はHMMであり、加法的歪みは式41のような単一のガウス密度としてモデル化される。これらの条件下で、微分により、次のような再推定公式が得られる。

【数53】

$$\mu'_{b_i} = \frac{\sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \gamma_t(n, m) E(b_{i,j} | y_{t,j}, s_t = n, c_t = m, \eta, \Lambda_X)}{\sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \gamma_t(n, m)} \quad (53)$$

【数54】

$$\sigma'^2_{b_i} = \frac{\sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \gamma_t(n, m) E(b_{i,j}^2 | y_{t,j}, s_t = n, c_t = m, \eta, \Lambda_X)}{\sum_{t=1}^T \sum_{n=1}^N \sum_{m=1}^M \gamma_t(n, m)} - \mu'^2_{b_i} \quad (54)$$

前のように、 $\gamma_t = p(Y, s_t = n, c_t = m | \eta, \Lambda_X)$ は、Yと、観測値y<sub>t</sub>を生成した変換済みモデル $\Lambda_Y = G_\eta(\Lambda_X)$ におけるn番目の状態の第mミクス★

☆チャとの結合尤度である。式53および式54における条件付き期待値は次のように評価される。

【数55】

$$E(b_{i,j} | y_{t,j}, s_t = n, c_t = m, \eta, \Lambda_X) = \mu_{b_i} + \frac{\sigma_{b_i}^2}{\sigma_{n,m,j}^2 + \sigma_{b_i}^2} (y_{t,j} - \mu_{n,m,j} - \mu_{b_i}) \quad (55)$$

【数56】

$$E(b_{i,j}^2 | y_{t,j}, s_t = n, c_t = m, \eta, \Lambda_X) = \frac{\sigma_{b_i}^2 \sigma_{n,m,j}^2}{\sigma_{b_i}^2 + \sigma_{n,m,j}^2} + \left\{ E(b_{i,j} | y_{t,j}, s_t = n, c_t = m, \eta, \Lambda_X) \right\}^2 \quad (56)$$

【0071】式55を調べると、EMアルゴリズムの収束性について観察することができる。 $\sigma_{b_i}^2$ が小さい場合、収束は遅い。これは、われわれの実験（第5節）の場合もそうであり、異なるトランスデューサおよび伝送

チャンネルによる不一致の分散が小さいためである。決定的バイアス（ $\sigma_{b_i}^2 = 0$ ）の極限の場合、推定値は全く変化しない。これは、式47を使用して $\mu_b$ を推定し式54を使用して $\sigma_b^2$ を推定することによって補正する

31

ことができる。

【0072】式34の加法的モデルのもとで特徴空間およびモデル空間におけるバイアスパラメータを推定する方法を示した。しかし、加法的バイアスモデルはケプストラム特徴にのみ適用されている。われわれの実験では、ケプストラム特徴に加えて、デルタおよびデルターデルタケプストラム特徴ならびにデルタおよびデルターデルタ対数エネルギー特徴を使用した。確率論的マッチングアルゴリズムでは、デルタおよびデルターデルタ対数エネルギー特徴は変換しない。しかし、デルタケプストラムおよびデルターデルタケプストラムに対する不一致の効果は考慮する。特徴空間バイアスモデルでは、デ

$$\Delta C_{l,m} = \sum_{k=-K}^K GkC_{l-k,m} \quad (57)$$

ただし、 $\Delta C_{l,m}$ および $C_{l,m}$ はそれぞれ、1番目の時間フレームに対するm番目のデルタケプストラム係数およびm番目のケプストラム係数である。Gは0.375に※

$$\Delta^2 C_{l,m} = \sum_{n=-N}^N Gn\Delta C_{l-n,m} \quad (58)$$

ただし、 $\Delta^2 C_{l,m}$ は1番目の時間フレームに対するm番目のデルターデルタケプストラム係数である。G=0.375およびN=1と選ぶ。異なるフレームに対するケプストラム係数は独立であると仮定すると、デルタケプ

$$\sigma_{\Delta C_{l,m}}^2 = \sum_{k=-K}^K G^2 k^2 \sigma_{C_{l-k,m}}^2 \quad (59)$$

ただし、 $\sigma_{\Delta C_{l,m}}^2$ および $\sigma_{C_{l,m}}^2$ は、1番目の時間フレームのデルタケプストラムおよびケプストラムの第m成分の分散である。同様に、デルターデルタケプストラム☆

$$\sigma_{\Delta^2 C_{l,m}}^2 = \sum_{n=-N}^N \sum_{k=-K}^K G^4 n^2 k^2 \sigma_{C_{l-k-n,m}}^2 \quad (60)$$

デルタおよびデルターデルタバイアス項の分散を推定することに興味がある。バイアスは分散が $\sigma_b^2$ のi.i.d. ガウシアンであると仮定されるので、デルタバイア

$$\sigma_{\Delta b_i}^2 = \sum_{k=-K}^K G^2 k^2 \sigma_{b_i}^2 \quad (61)$$

同様に、デルターデルタバイアスの第i成分の分散は次のように推定することができる。

$$\sigma_{\Delta^2 b_i}^2 = \sum_{n=-N}^N \sum_{k=-K}^K G^4 n^2 k^2 \sigma_{b_i}^2 \quad (62)$$

【0073】観察されるように、歪みに使用される統計モデルは単純なガウス密度である。上記と同じ確率論的マッチングアルゴリズムは、より一般的なモデル変換の場合にも適用可能である。

【0074】

32

＊ルタおよびデルターデルタケプストラム特徴は不一致によって影響を受けないと仮定する。すなわち、デルタおよびデルターデルタバイアスベクトルは0であると仮定する。これは、ケプストラムのバイアスが発声全体に対して一定であると仮定する場合には意味のある仮定である。同様に、モデル空間では、デルタおよびデルターデルタ平均値ベクトルが0であると仮定する。しかし、デルタおよびデルターデルタ分散については仮定しない。これらの分散ベクトルは以下のように推定する。デルタケプストラムは次式に従って計算される。

【数57】

※固定された利得項であり、K=2である。デルターデルタケプストラムは次式に従って計算される。

【数58】

★ストラムの分散は、ケプストラムの分散を用いて次のように書くことができる。

【数59】

☆の分散を次のように導出することができる。

【数60】

◆スの第i成分の分散は式59を使用して次のように推定することができる。

【数61】

＊【数62】

【発明の効果】以上述べたごとく、本発明によれば、本発明に従って形成したシステムは、個々のアプリケーションごとにトレーニングする必要はないという効果がある。また、本発明に従って形成したシステムは、プールモデルやコードブックを使用する既知のシステムよりも

高い精度を有するという効果がある。

【図面の簡単な説明】

【図 1】 不一致問題を解決する既知のシステムの図である。

【図 2】 本発明に従って形成したシステムの第 1 実施例の図である。

【図 3】 本発明に従って形成したシステムの第 2 実施例の図である。

【図 4】 トレーニングと試験の不一致を示す図である。

【図 5】 式 7 の結合最大化を示す図である。

【図 6】 式 7 の結合最大化を示す図である。

【符号の説明】

100 変換器

102 オフセット計算器

104 コードブック

106 比較器

108 HMM記憶装置

200 変換器

202 変換プロセッサ

204 変換記憶装置

206 比較器

208 HMM記憶装置

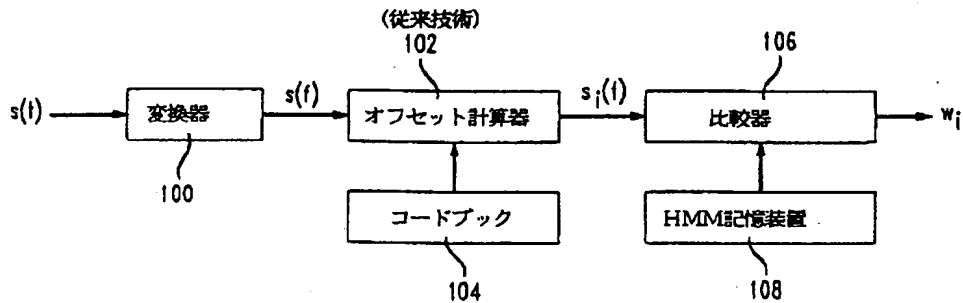
10 300 変換器

302 比較器

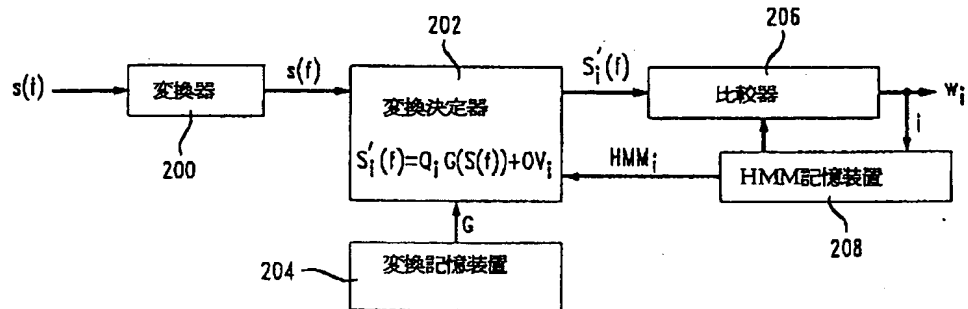
304 HMM記憶装置

306 変換プロセッサ

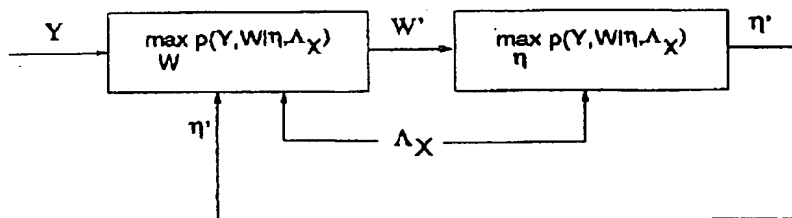
【図 1】



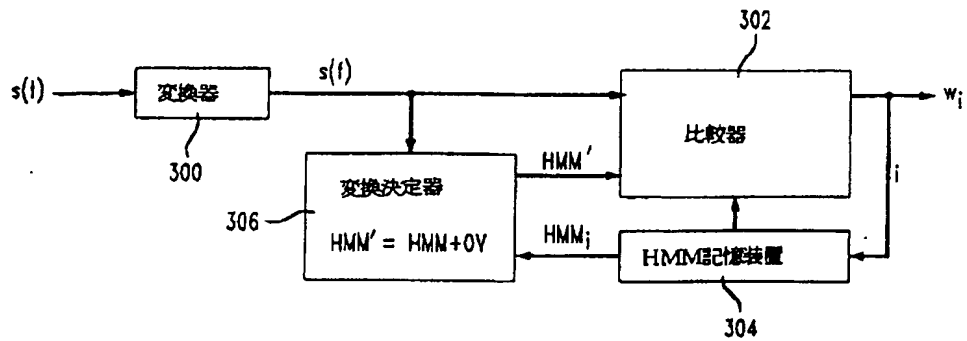
【図 2】



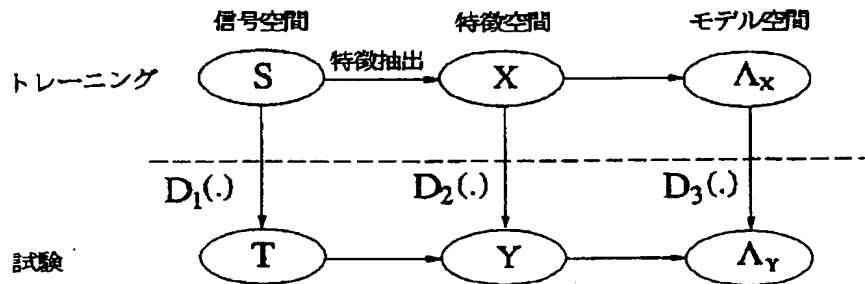
【図 6】



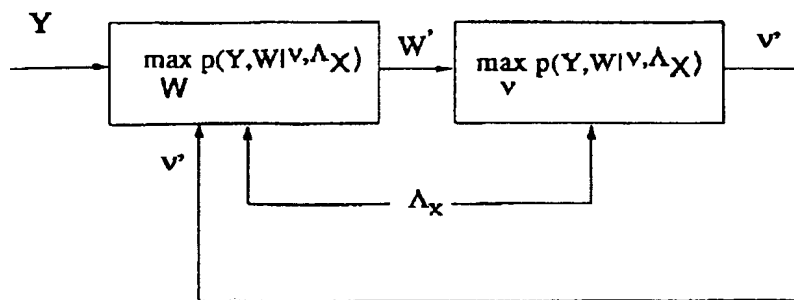
【図 3】



【図 4】



【図 5】



フロントページの続き

(72)発明者 アナンス サンカー  
 アメリカ合衆国, 94555 カリフォルニア,  
 フレモント, トウペロ ストリート  
 34367